

Bettina Zurstrassen

A Sceptical Look at the Quantitative Education Research in Civic and Citizenship Education

At the latest since the PISA study, the international comparability of educational system performance is called for. Little discussion takes place on the research problems occurring in the course of internationally comparative educational systems in the social societal teaching subjects. Research, however, finds itself confronted with methodical, scientific-theoretic and test-ethical problems. These problem areas are to be outlined in this paper. The focus will be on quantitatively comparative education studies. Essentially the analysis consists of an examination of the ICCS Study (International Civic and Citizenship Education Study) of the biggest internationally comparative political education study in existence.

Keywords

Civic and citizenship education, test ethics, Gender

Introduction

At the latest since the PISA study, the international comparability of educational system performance is called for. Little discussion takes place on the research problems occurring in the course of internationally comparative educational systems in the social societal teaching subjects.

It is true that the initial, if only tentative, steps have been taken for several years towards (international) comparative quantitative education research in the field of civic and citizen education. Research, however, finds itself confronted with methodical, scientific-theoretic and test-ethical problems, whose range has until now only been discussed “superficially” (cf. Hahn 2010, Schulz 2009, 115).

These problem areas are to be outlined in this paper. The focus will be on quantitatively comparative education studies. Essentially the analysis consists of an examination of the ICCS Study (International Civic and Citizenship Education Study) of the biggest internationally comparative political education study in existence.

Four aspects are thereby of central significance:

1. The problem of particularism in civic and citizenship education.
2. There is no existing educational-theoretic and empirically verified model of proficiency, suitable internationally for making deductions.
- 2.1. The dilemma of domain-specific diagnostics.
3. Scientific-theoretical problems of interdisciplinary-oriented civic and citizenship education research.
4. Test ethics for quantitative education studies in the field of civic and citizenship education.

Regarding the statistically methodical debate being held in comparative social research and comparative psychology, I shall only treat this in the paper in a cursory manner.

1 Because of the negative experience in Nazi Germany, education and cultural policy is subject to regulation by the individual federal states. This means to avoid enforced conformity of the education system

1. The Problem of Particularism in Civic and Citizenship Education

In civic and citizenship education considerable particularism exists. The subject structure, the curricula and the concepts for civic and citizenship education vary considerably in individual national and federal states. In the Federal Republic of Germany, by reason of cultural sovereignty¹, we can even consider it a patchwork system, since each of the 16 federal states has its own curriculum, which partially vary in objectives and choice of content considerably. The problem of international particularism in civic and citizenship education is extensively thematised in the ICCS study and in the Eurydice study² and empirically reviewed (cf. Schulz et al. 2010a; 2010b; Eurydice 2005). The aspect of particularism in civic and citizenship education is therefore presented only superficially in this article.

1.1. Particular Subject Structure and curricular Particularism

Contrary to the teaching subjects of mathematics or teaching the country's language, in many states there is no teaching subject distinctly for civic and citizenship education. In the 38 countries that took part in the ICCS (International Civic and Citizenship Education Study), for instance, there are three organizational forms at the curricular level for lower secondary education: 1. Civic and citizenship education as a specific, stand-alone subject (either compulsory or optional). 2. Civic and citizenship education integrated into other subjects. 3. Civic and citizenship education as a cross-curricular theme” (cf. Schulz et al. 2010 a, 22). Many countries use two or three of these curricular approaches simultaneously. In 18 of the 38 participating countries, there is no stand-alone subject known as ‘Civic Education’ (ibid., 23). This particularisation is further differentiated due to the wide thematic distribution in the national syllabuses and the fact that in those states which have their own teaching subject for civic education, the

2 The position “civic education” in the member states was examined in the Eurydice Study by order of the European Union.



subject is often not taught in all class grades (cf. *ibid.*; Eurydice 2005, 20 a). The test contents of the ICCS study are therefore often not coextensive with the contents actually taught. The deficient curricular validity warns us to be careful when deriving judgements from the study to apply them to the teaching quality. In order to enhance the curricular validity, the ICCS research group developed regional modules for the survey in 2008/09 (cf. Schulz et al. 2010, 15).

1.2 Competing Programmatic Subject Concepts?

The ICCS study differentiates between two subject-didactical, conceptual approaches which are described as follows: "Civic education focuses on knowledge and understanding of formal institutions and processes of civic life (such as voting in elections)" (cf. Schulz et al. 2010 a, 14). Conversely, the concept of citizenship education focuses on "knowledge and understanding and on opportunities for participation and engagement in both civic and society" (*ibid.*). This differentiation is superficial. The theoretical concepts and debates about Civic Education, Citizenship Education, Political Education and Democratic Education vary internationally. Supporters of civic education will complain that the practice of political active skills is not a stand-alone characteristic of citizenship education. No empirically valid data are available which permit serious statements of how far the various conceptional notions affect the specific lesson and learning success of the students. Due to the limited page number the debate will not be reintroduced (Georgi, 2008; May 2008). The debate is chiefly concerned with educational and research-political questions of authority, since research funding, personnel resources and definition responsibility are involved in the denomination of professorship appointments.

2. No Education-Theoretical and Empirically Proven, Internationally Adaptable Proficiency Model

The circumstances described above complicate the development of an internationally adaptable education test. They are complicate above all because no internationally adaptable education-theoretically founded and empirically proven proficiency model for civic and citizenship education is available (cf. Himmelmann 2005).

If we look at the test item for the ICCS Study 2009 (Schulz et al. 2010 b), we see a melange of topic complexes which are questioned. There is, however, no subject-didactical and education-theoretical foundation for the knowledge test. It is not clear on which proficiency model the knowledge test of the ICCS study is based. A proficiency model giving education-theoretical accountability for which skills and special

knowledge the student should possess in order to become politically and socially capable of action is not available for the research group. Not even specialised didactic categories for the choice of test items are defined. Instead we have the choice based on specialised categories: „Civic society and systems“, „civic principles“, „civic participation“, „civic identities“ (*ibid.*, 60). The research group of the ICCS study defines three proficiency levels (*ibid.* 61); their empirical discrimination and verification, however, remain unclear. The result is that the entire knowledge test floats in a vacuum, above all since the educational relevance of the test item is unclear. To be fair, however, it must be emphasised that the domain-specific diagnostic is still in the beginning stages. Based on the present status of research, central target setting for civic and citizenship education by means of quantitative test procedures cannot be empirically realised, if indeed they can be adopted at all for quantitative studies. This is exemplified below, demonstrated by the example "political judgement capability".

2.1 The Dilemma of Domain Specific Diagnostics

In the course of the ISSC study, 63,000 teachers from 38 states taking part were asked "how they conceptualized civic and citizenship education, what they saw as objectives, and how this subject area was being delivered in their school" (cf. Schulz et al. 2010a, 3, 64). From 10 given categories of "civic and citizenship education study" they were required to select the three most important ones. One of these categories was "promoting students' critical and independent thinking", which was defined by teachers from Cyprus, Finland, Latvia, Liechtenstein, Lithuania, Slovenia, Spain, Sweden as most important aim (*ibid.* 68). This item represents the proficiency field "political judgement capability" for civic and citizenship education also in Germany.

This article utilises the following definition of the term 'political judgement': Political judgements are based on expertise. This distinguishes them from prejudices and anticipatory judgements. They are developed in a process of consideration and are justifiable; categories of value rationality and purposive rationality are enlisted in the process. They can be elaborated through dialogue. Makers of judgements are aware of the perspective nature of their own positions. Political judgements must be accessible to reflection (cf. Berliner 2009; Massing 2006, 157; Weinbrenner 1997, 75 ff; Harwood, Hahn 1990). The emphasis of presentation is upon the problems that arise in quantitative methods for the measurement of judgement competence within the framework of achievement tests. Because teaching subjects in the social sciences are among the interpretive subjects, the domain-specific diagnostic approach works chiefly with qualitative

methods in empirical social research. Given the complex structure of the process whereby political judgements are formed, some experts of teaching methodologies take a sceptical view of approaches that attempt to identify political judgement competence on the basis of diagnostic testing. Weinbrenner points to the model and variable formation required if the process of judgement formation is to be discussable at all from a research point of view (cf. Weinbrenner 1997, 81). Along with theories of developmental and cognitive psychology, efforts to identify judgement competence must also take the structural characteristics of the discipline into account. In this article, the problem can only be pointed up by way of example: In a test designed to identify the structures of political judgements, respondents are asked the following question, for instance: 'What could the consequences of more direct democracy be?' Respondents may tick these answer categories:

- a) Direct democracy can help remedy the overwhelming power of the parties.
- b) Direct democracy can reduce disenchantment with politics.
- c) Direct democracy is cheaper.
- d) More democracy can lead to decisions based on political mood.
- e) The people cannot cope with difficult decisions (Massing, Schattschneider 2005, 23).

There are arguments for and against all five answer categories. Even answer category 'c' can be justified in reference to the costs of citizens' protests, the many years consumed by proceedings before administrative law courts or the macroeconomic cost of increases in acts of violence by political-extremist groups as a result of disenchantment with politics. Because all of the answer categories are discursive in nature, it is unclear just how one should decide which criteria should be viewed as companion to a higher degree of sophistication of political judgement competence. If the political judgement of a test's designers were set as the applicable standard, this would represent a violation of the requirement that empirical quality criteria must be objective. The diagnostic identification of political judgement competence confronts the didactics of civic and citizenship education with additional problems, which for the most part are not limited to quantitative survey methods:

- There is no appropriate definition of what constitutes a political judgement, and of the characteristics inherent to a political judgement (cf. Berliner 2006; Bourke, Geuss 2006; Weinbrenner 1997, 73 f)
- The process of human development towards the formation of judgements and the stages of this development are, despite the structuralist-cognitive

theory of development of Jean Piaget and Kohlberg's stage model of moral development and recent studies in developmental research (Colby, Kohlberg 1987; Fend et al. 2009), still inadequately explained, both empirically and theoretically.

- Because political judgements are syntheses of judgements of fact and judgements of value, they cannot be assigned to dichotomous answer categories such as 'right' and 'wrong' (see the test item shown above).
- Political topics have different degrees of affective concern, and this could have an impact on the sophistication of the political judgement form. As a result, a multiplicity of measurement values is required to neutralise this factor of influence mathematically.
- There are no clearly defined principles for the construction of political judgements (of the form seen, for instance in Latin instruction for the construction of the gerund). Test items for reliability review or repeat measurements cannot be reproduced merely by making slight variations in the items themselves (e.g. modifying numerical values).

What We Need to Do

Before the domain enters into quantitative learning status research, we need an internationally adaptable, educational-theoretically based proficiency model be conceptualised by an internationally composed research team. In the second phase, test items and diagnostic instruments for checking the proficiency model must be developed and tested. Only then, based on this, can a test for internationally comparative education studies (learning status investigation) be constructed and implemented. Implementation of such a procedure could take a long time to develop and use up financial resources.

3. Scientific Theoretical Problems of Interdisciplinary Civic and Citizenship Education Research

The question of how democratic proficiency can be acquired and how political judgement and activity proficiency develop are of vital significance for democratic systems. Answers to these questions are being sought by various specialised disciplines, for example by psychology, educational sciences, political ethnology, political philosophy, political science and political didactics. They all examine sections of the complex research material from their own specific expertise perspectives. With regard to the research fields, there is considerable differentiation, e.g. political socialisation research, political attitude research, political culture research, research on teaching of a specific expertise, research in the teaching profession, learning diagnostics as well as curriculum and school book research, whereby the list could easily be ex-



tended. Political education and socialisation research is a domain – and here it differs from many other teaching subjects – which cannot be classified into only one subject discipline. Interdisciplinary cooperation as applied to international research projects is almost unavoidable in order to cope with the complexity of the research topic. Furthermore, it must be taken into account that test values from internationally comparative education studies can be distorted, since participants of the test interpret the items according to their own culture. The conceptuality, the concepts and the feasibility of interdisciplinary research are contentious. In the scientific theoretical debate there is, in the meanwhile, an extensive and long-lasting discussion on the concepts of what is multi-, inter- and transdisciplinary (cf. Jantsch 1970, Heckhausen 1972, Klein 1990; Moran 2002; Mittelstraß 2003; Jungert et. al 2010), which cannot be comprehensively presented here. Heckhausen differentiates in his category system between six grades of interdisciplinarity. These range from “indiscriminate interdisciplinarity”, i.e. various teaching subjects are presented “adjacent to each other” up to “unifying interdisciplinarity”, whereby a merging of theory approaches and methods of different subjects takes place in the research process (cf. Heckhausen 1972, 87-89).

The requirement of “interdisciplinarity” is easier said than done. Interdisciplinary research in the field of civic and citizenship education is confronted with a number of problems which must be reflected in the research process. Firstly, a variety of faculty cultures exists, e.g. disciplinary discussion cultures, paradigms, publication strategies (c.f. Surkopp 2010, 14). Secondly, in civic and citizenship education, various programmatic conceptual approaches exist with regard to content, methods and objectives of the particular subject (see Chap. 1) and thirdly, various expert scientific theories and terminology concepts exist. The latter aspect will be considered below.

3.1 Various Scientific Disciplinary Terminology and Theory Concepts

Heckhausen measures the degree of interdisciplinarity based on indicators of theoretical integration. The problem is illustrated through examples of use of terminology and theory concepts. Behind scientific specialised terminology there are theory concepts which are interpreted differently, e.g. terms for politics, terms for democracy, terms for justice, the controversy of a narrow or wider term for violence in sociology, psychology and education science or the term “environmental awareness” (c.f. Rippel 2004, 16). The “battle of words” is waged within disciplines as well as interdisciplinarily.

The use of terminology has an impact on research results. The degree of women’s political activation is,

for example, within the narrow political sense often meagre, whereas when a wide political term is applied it takes on more significance (see Chap. 4). The fact that for the subject matter of an examination often extremely heterogenic data exist is often explained by the different applications of the terminology concepts.

What We Need to Do

The item development for education studies is guided in theory as a rule by education studies. In order to avoid severe distortion of the test values (see chapter 3, 4), terminology and theory concepts in the case of interdisciplinary education studies, when optimising the items, must be made known and coordinated. Furthermore the research design, the survey method and the individual steps of the research process must be made transparent, to provide sequential compliance of the research process and its results among the personnel.

4. Test Ethics in the Case of Quantitative Education Studies in the Field of Civic and Citizenship Education

The aspect of test ethics for education studies in civic and citizenship education has been little thematised until now. In the USA there is, in the meanwhile, a wide debate on the research methods used for test ethics (cf. www.fairtest.org; Nichols, Berliner 2007; McNeil 2001; Sacks 2000; Gipps, Murphy 1994). Test ethics in education studies examine to what extent those questioned who were from different social groups and cultural socialisation have unequal chances of success because of the test construction. It examines, furthermore, the problems of normative overcoming by item formulation and the operationalisation of answer categories. In this chapter, three aspects of test ethics are outlined which possess great relevance in quantitative education studies of civic and citizenship education: “political culture and test ethics”, “social structure and test ethics” and thirdly, “political-didactical test ethics”.

4.1 Political Culture and Test Ethics

In comparative social scientific culture research and in comparative psychology there is a wide, statistical discussion on methods regarding the comparability of data expressed in terms of the concept of equivalence. Van de Vijver defines the term “equivalence” as follows: “Equivalence refers to the question whether there is any difference in measurement level of within- and between-group comparisons. If the measure is biased against some cultural group, individual differences within a cultural population and across cultural populations are not measured at the same scale” (1998, 43). Contrary to education research in mathematics, social scientific education research must

examine very closely the culturally specific dimension of social structures and terminology in order to avoid measurement errors. Van der Vijver and Hambleton differentiate between three distinct types of bias that may affect the validity of tests that have been adapted for different cultural contexts:

Construct bias: A measuring tool is made operational for a survey group, e.g. the construct "Attitude towards the social state." Using this measuring tool in a different culture can lead to a distortion of data, if the construct is defined otherwise in this culture.

Item bias: Distortion of compiled data due to inaccurate translation. But even in the correct translation distortion can arise if, for example, within one culture there are different reactions on the answer scale.

Method bias: In one culture, individuals react systematically to a survey instrument in a different way to those questioned from a different culture (cf. Van der Vijver, Hambleton 1999, 89-99).

In the ICCS study, politically cultural differences, above all in the context of methodological discussion (e.g. translation and re-translation of the questionnaire, national pre-tests) are thematised. Furthermore, social structural aspects are touched on which lead to unequal chances of success in the education test. A discussion with culturally different interpretation patterns of social and political terminology and theory constructs does not take place. Furthermore, "culture specific" political behaviour patterns such as Carol Hahn presents in her study are not thematised. Hahn, Professor of Education at Emory University, performed ethnologically oriented observations of teaching in six industrialised countries (Great Britain, the US, Germany, Denmark, Netherlands and Australia) (Hahn 1999). She verified her observations through interviews with pupils and instructors, and through a quantitative survey. Hahn describes the classroom teaching and the impact it has on the pupils' patterns of political attitudes (ibid. 241). While 76 percent of the Danish pupils surveyed expressed a high willingness to discuss politics with other people, this willingness was echoed by only 23 percent of Dutch participants in the survey. Hahn attributes the difference to civics instruction in Denmark, where a highly developed culture of political discussion was cultivated. In the Netherlands, she observed, where there is no self-contained instruction in 'Citizenship Education', these experiences are missing (ibid. 243; see also Maiello et al. 2003, 393). Her analysis is focused upon the culture of instruction. Hahn does not attempt to embed the results of her research in the political culture of the individual countries examined. This would be required for international comparative civic and citizenship educational studies of educational achievement. How does political culture affect response behaviour if, for instance, a study is to formulate political judgements even though in a par-

ticular participant country it is taboo to discuss one's political opinion in a public setting?

4.2 Social Structure and Test Ethics

For many years, there has been a debate in empirical methodology about the language style used, which in surveys and tests is often middle-class-oriented (cf. www.fairtest.org). When taking performance tests, pupils from 'language-poor' families or pupils from families with a migration background thus must not only demonstrate their expertise in the subject but also encode the test items in their own language. Conceivably, some test items will go unanswered for reasons of language hurdles – even though the relevant expertise exists (cf. Bartnitzky 2008). This problem must also be taken into consideration in tests in civic education.

4.2.1 Gender and Test Design

Is there a gender gap in political specialised knowledge? The empirical data status is considerably more heterogenic than scientific discourse would have us believe, which is aimed at discrimination against women. Whereas Wolak and McDevitt in their study detect a lower political specialised knowledge among young women (2010, 2), girls in the ICCS study in the vast majority of the states involved (31) on average achieved considerably higher test values than young men (cf. Schulz et al. 2010a, 39-40). The long term comparison of the ICCS study (including the CIVIC study) indicates that the female respondents in the surveys of 1971 and 1999 came off even worse than the male respondents, but the values have progressively converged (cf. Schulz et al. 2010b, 80). How can the heterogenic data status be explained? An analysis of the test designs is unavoidable. This, however, is conditional to the test tools being accessible for the public, something, however, which is often not the case. A positive example is given by Wolak and McDevitt, who have listed the test items in the attachment to their article (Wolak, McDevitt 2010). Based on analysis of their survey instrument, an attempt at explanation for the poor results from women in their study can be developed. The authors explain their research results with the „theory of risk aversion“, which implies that women mark the answer category "don't know", whereas, if men do not know a result will make a guess (ibid. 3). Lizotte and Sidman arrive at similar results in their study: "We argue that it is this propensity to guess under conditions of uncertainty that consistently produces political knowledge scales that underestimate the ability of women" (2009, 128). There is no category "don't know" in the ICCS study. It should be examined how far the construction of the item and answer categories distort the test value. It must be examined in general, to what extent the test construction generates gender-specific differences.

Important results concerning test construction and the gender gap in political knowledge are discussed in the article "Stereotype Threat and the Gender Gap in Political Knowledge" by McGlone, Aromson and Kobrynowitz: "These results show that the gender gap in political knowledge narrowed substantially when two potentially stereotype-threatening elements of the survey context were eliminated. Specifically, when women (rather than men) were interviewed by female interviewers and the survey was portrayed as nondiagnostic (rather than diagnostic) of alleged gender differences in political knowledge, the gender gap closed (2006, 396)."

A further attempt at explanation exists in the range of topics of the questioned test knowledge. The political fields of interest of women and men are different. Whereas women are interested primarily in the politics of education, social matters and family, men show greater interest in the fields of party politics, societal institutions as well as business and foreign policy (cf. Noelle-Neumann, Köcher 2002, 395 f.). Wolak and McDevitt, however, in their study asked knowledge questions about party politics throughout. Setting a broad or narrow political terminology for the content choice of test items can lead to a distortion of the result valuation. I would like to formulate this here as a hypothesis. An investigation of this hypothesis calls for a more in-depth evaluation of the selection of items in previous studies of achievements in civic education; this evaluation is still in progress.

4.3 Political-Didactical Test Ethics

The prohibition of political indoctrination and all other indoctrination on the one hand and the decree of objectivity on the other hand are in many states central paradigms of civic and citizenship education. Infraction of these political didactical principles in quantitative studies is to be exemplified by an illustration on the basis of a test item from the ICCS study. I shall merely formulate theses which would need to be confirmed through systematic analysis of quantitative test material for the domain. The 14-year-old pupils taking part are to resolve the following task: A pupil has bought new shoes; then he discovers that these "were made by a company that employs young children to make the shoes in a factory and pays them very little money for their work" (Schulz et al. 2010b, 67). The respondent is to mark how the pupil can convince other consumers not to buy shoes when the production involves child workers. The following answer categories are offered:

How can he best try to do this?

- Buy all of the shoes himself so no one else can buy them
- Return the shoes to the shop and ask for his money back
- Block the entrance to the shop so people cannot enter it

- Inform other people about how shoes are made
- The last answer is considered to be the right answer. From a political didactical viewpoint, this example illustrates the following ethical conflict situations

Necessitation of Oversimplified Political Judgements

Ostensibly open explanation of child labour and the suggested boycott of products in this category receive the widest political acceptance. In social science research, the boycott of goods produced by children, however, is seen in the meanwhile in very differing ways (cf. Liebel 2001, 9). The morally congenial behaviour of boycotting the goods can be contra-productive for the child workers. By boycotting the goods the child workers lose their essential source of income, if simultaneously compensation measures are not taken. In particular in national economies, where child workers are a significant economic factor, wage levels of the parents must be increased, orphans would need to receive transfer benefits, so that they don't need to work. Consumer behaviour is of considerable relevance in the industrialised states. Are they – including the test participants – willing to pay higher prices for goods in order that the wage levels in the so-called emerging countries and the working conditions can be improved? The test respondents are not given the chance of developing alternative solution strategies. My first hypothesis is: In the case of quantitative education studies in the field of "civic and citizenship education" the test respondents are forced to make oversimplified judgements of a political and economic nature.

Adoption of the Valuation Structure of the Test Designers

The sociologist Liebel discusses in his publication also the positive effects of child labour, since this is connected to the chance of improving the standard of living or gaining professional experience (ibid.). His explanations are not a plea for exploitation of child labour, yet it becomes clear that the normative borders of interpretation are outside those given in the test. The test respondents are not, however, given the chance to formulate their own normative judgement. Instead, the survey respondents in quantitative education studies on civic and citizenship education, according to my second hypothesis, have the valuation structure of the test designers forced upon them. In order to conclude the test successfully, the respondents have to try and follow and adopt the interpreting pattern of the test designers.

Definition Sovereignty over the Concept of the Politically Active Citizen

The test item presented above describes four political courses of action. „Inform other people about how

shoes are made" is defined as the correct answer. With this, however, a high standard of political, responsibly active citizen is formulated, which assumes a very high level of participation. Other forms of political protest are defined as "wrong" (see answer categories 1, 2 and 3) or are not offered. Many consumers will not accept that the answer category "Return the shoes to the shop and ask for his money back" as being wrong, since in this way they are making use of their power as consumers. The political activation level and the political scope with regard to educating other consumers is indeed less than with answer "d", but it is not wrong. The test designers, according to my third hypothesis, force their interpretation pattern of the politically responsible, active citizens and legitimate political forms of action onto the test participants. These problems could occur systematically, especially in test items which are aimed at testing action proficiency.

What We Need to Do

In civic and citizenship education, there is a necessity for more intensive research and more intense sensitisation for "test design errors" which lead to a systematic, socially structured, cultural and normative distortion of test results. This applies not only with respect to the design of survey instruments, the content choice for the education tests, but also for interpretation of the data.

Conclusion

Education tests have a considerable scandalisation as well as a dramatisation potential and therefore influence political performers greatly. Social reality is designed on the basis of empirical data. The danger of

instrumentalisation of education studies by imposing political aims is immense. Critical analysis of the empirical data in science and politics is consequently fundamental. I have illustrated in the article that in the case of quantitative education studies in civic and citizenship education, significant distortion can occur which arise from the complexity and distinctiveness of the research field. These aspects were analysed superficially in the article. Further, in-depth examinations of the individual, partial aspects are required. Quantitative political education research must be unpretentious, since:

1. No education-theoretical and empirically assured proficiency model for civic and citizenship education exists which is internationally adoptable.
2. Even at the level of testing specialised knowledge, distortion of valuation can occur, due to variations in culturally specific, socially structural or interdisciplinary interpretations of terminology and theory concepts.
3. Numerous aims of civic and citizenship education, by reason of their complexity and normativity, mean that quantitative surveys cannot provide the necessary ability to operate. Thus, for example, the educational aim of "promoting students' critical and independent thinking", is beyond quantitative measurability, since in education tests, no scientifically objective categories of "right answer" or "wrong answer" can be defined.

My discourse in this article does not oppose comparative (international) quantitative education research in principle, but it points out the research problems and advises caution against unconsidered handling of research results.

References

- Abs, Josef, eds. 2009. *Introducing Quality Assurance of Education for Democratic Citizenship in Schools*. Strasbourg.
- Bartnitzky, Horst. 2008. VERA Deutsch 2008: Wie Kinder durch Vergleichsarbeiten Unrecht geschieht [How Children are wronged by Comparative Tests]. In: Grundsschulverband e.V. (http://www.grundsschulverband.de/fileadmin/Bildungspolitik/GS_103.Bartnitzky_Vera08_Deutsch.pdf).
- Berger, Guy. 1972. Introduction. In: Apostel, Léo; Berger, Guy; Briggs, Asa; Michaud, Guy, eds. *Interdisciplinarity. Problems of Teaching and Research in Universities*. Paris, 23–26.
- Berliner, Isaiah. 2006. *On Political Judgement*. New York.
- Böhm-Kasper, Oliver. 2006. Politische Partizipation von Jugendlichen. Der Einfluss von Gleichaltrigen, Familie und Schule auf die politische Teilhabe Heranwachsender [Political Participation of Adolescents. The Influence of Peers, Family and School on Youths' Political Participation]. In: Helsper, Werner u.a., eds. *Unpolitische Jugend? Eine Studie zum Verhältnis von Schule, Anerkennung und Politik [Unpolitical Youth? An Analysis of the Relation between School, Recognition and Politics]*. Wiesbaden, 53–74.
- Bourke, Richard; Geuss, Raymond, eds. 2009. *Political Judgement*. Cambridge.
- Bridges, David; Smeyers, Paul; Smith, Richard, eds. 2009. *Evidence-Based Education Policy*. Oxford.
- Büner, Markus. 2006. Einführung in die Test- und Fragebogenkonstruktion [Introduction to the Construction of Testing and Questionnaire]. München.
- Center for Civics Education. 2010. National Standards for Civic and Government. In: Center for Civics Education. (http://new.civiced.org/images/stories/PDFs/Publications/National_Standards/National_Standards_full_text.pdf).
- Colby, Anne; Kohlberg, Lawrence. 1987. *The Measurement of Moral Judgment. Volume I: Theoretical Foundations and Research Validation*. Cambridge.
- Colby, Anne; Kohlberg, Lawrence. 1987. *The Measurement of Moral Judgment. Volume II: Standard Issue Scoring Manual*. Cambridge.
- Dünnebier, Katrin; Gräsel, Cornelia; Krolak-Schwerdt, Sabine. 2009. Urteilsverzerrung in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten [Bias of Assessment in Educational Performance Appraisal. An Experimental Study about Anchor Effects]. In: *Zeitschrift für Pädagogische Psychologie*, Jg. 23 (3-4), 187–195.
- Eurydice 2005. *Citizenship Education at School in Europe*. (<http://www.moec.gov.cy/programs/eurydice/publication.pdf>).
- Farnen, Russel; German, Daniel; Dekker, Henk; De Landtsheer, Christl; Suenker, Hein. 2008. *Political Culture, Socialization, Democracy and Education*. Frankfurt.
- Fend, Helmut; Berger, Fred; Grob, Urs, eds. 2009. *Lebensläufe, Lebensbewältigung, Lebensglück. Ergebnisse der LIFE-Studie [Lifecourses, Coping with Life, Fortune. Results of the LIFE-Study]*. Wiesbaden.
- Fuchs, Gesine. 2008. Politik: verfasste politische Partizipation von Frauen [Politics: Constituted Political Participation of Women]. In: Becker, Ruth; Kortendieck, Beate, eds. *Handbuch Frauen- und Geschlechterforschung. Theorie, Methoden, Empirie [Compendium Women and Gender Studies. Theory, Methods, Empiricism]*. Wiesbaden, 539–546.
- Georgi, Viola B. 2008. *The Making of Citizenship in Europe: New Perspectives on Citizenship Education*. Bonn.
- Gill, Jungil. 2008. Gender Differences in a Pre-adult Learning of Participatory Orientations: Comparison of the US, Korea, and Japan. In: Farnen, Russell et al. eds. *Political Culture, Socialization, Democracy, and Education*. Frankfurt, 181–198.
- Gipps, Caroline; Murphy, Patricia, 1999. *A Fair Test: Assessment, Achievement and Equity*. Buckingham.
- Hahn, Carol L. 1999. Citizenship Education: An empirical study of policy, practices and outcomes. In: *Oxford Review of Education*, Vol. 25 (1/2), 231–250.
- Hahn, Carole. 2010. Comparative Civic Education Research: What we know and what we need to do. In: *Citizenship Teaching and Learning*, Vol. 6 (1), 5–23.
- Haste, Helen. 2007. What Is a “Competent Citizen” and How Do We Create Them? In: Biedermann, Horst; Oser, Fritz; Quessel, Carsten, eds. *Vom Gelingen und Scheitern Politischer Bildung. Studien und Entwürfe [About Success and Failure of Civic Education. Studies and Designs]*, 75–96.
- Harwood, Angela M.; Hahn, Carole L. 1990. *Controversial Issues in the Classroom*. ERIC Clearinghouse for Social Studies/Social Science Education Bloomington IN.
- Heckhausen, Heinz. 1972. “Discipline and Interdisciplinarity”. In: Apostel, Léo; Berger, Guy; Briggs, Asa; Michaud,

- Guy, eds. *Problems of Teaching and Research in Universities*. Paris, 83–89.
- Helmke, Andreas. 2007. Unterrichtsqualität erfassen, bewerten, verbessern [Diagnosing, Assessing and Improving Tuitional Quality]. Stuttgart.
- Himmelmann, Gerhard. 2005. Was ist Demokratiekompetenz? [What is Democracy Competence?] In: <http://blk-demokratie.de/fileadmin/public/dokumente/Himmelmann2.pdf>
- International Test Commission. 2010. Guidelines for Translating and Adapting Tests. (<http://www.intestcom.org/upload/sitefiles/40.pdf>).
- Jungert, Michael; Romfeld, Elsa; Sukopp, Thomas; Voigt, Uwe, eds. 2010. Interdisziplinarität. Theorie, Praxis, Probleme [Interdisciplinarity. Theory, Practice, Problems]. Darmstadt.
- Klein, Julie Thompson. 1990. *Interdisciplinarity. History, Theory and Practice*. Detroit.
- Langner, Frank. 2007. Diagnostik als Herausforderung für die Politikdidaktik [Diagnostics as an Challenge for Political Didactic]. In: Schattschneider, Jessica, eds. *Domänenspezifische Diagnostik. Wissenschaftliche Beiträge für die politische Bildung* [Domain-Specific Diagnostics. Scientific Contributions to Education]. Schwalbach/Taunus, 58–70.
- Leven, Ingo; Quenzel, Gudrun; Hurrelmann, Klaus. 2010. Familie, Schule, Freizeit: Kontinuität im Wandel [Family, School, Recreation: Continuum in Transition]. In: Shell Deutschland Holding, eds. *Jugend 2010. Eine pragmatische Generation behauptet sich* [A Pragmatic Generation comes to stay]. Frankfurt am Main, 53–128.
- Liebel, Manfred. 2001. *Kindheit und Arbeit* [Childhood and Labour]. Frankfurt.
- Lizotte, Kate-Mary; Sidman, H. Andrew. 2009. Explaining the Gender Gap in Political Knowledge. In: *Politics & Gender*, Vol. 5, 127–151.
- Logan, Sarah; Medford, Emma. 2010. Gender Differences in the Strength of Association between Motivation, Competency Beliefs and Reading Skills. In: *Educational Research*, Vol 53, 85–94.
- Lutter, Andreas. 2011. Methoden zur didaktischen Rekonstruktion der Konfliktlösungs- und Urteilskompetenz. [Methods for Didactic Reconstruction of the Competence in Conflict Resolution and in Decision]. In: Zurstrassen, Bettina, eds. *Was passiert im Klassenzimmer? Methoden zur Evaluation, Diagnostik und Erforschung des sozialwissenschaftlichen Unterrichts* [What does happen in Classroom? Methods for Evaluating, Diagnostics and Investigation of Social Science Lessons]. Schwalbach/Taunus.
- MacBeath, John; Moos, Leif, eds. 2004. *Democratic Learning. The Challenge to School Effectiveness*. London.
- Maiello, Camine; Oser, Fritz; Biedermann, Horst. 2003. Civic Knowledge, Civic Skills and Civic Engagement. In: *European Educational Research Journal*, Vol. 2 (3), 384–395.
- Massing, Peter; Schattschneider, Jessica Simone. 2005. Aufgaben zu den Standards der politischen Bildung. Ergebnisse einer Pilotstudie [Tasks for Standards in Political Education. Results from a Pilot Study]. In: GPJE, eds. *Testaufgaben und Evaluation in der politischen Bildung* [Test Tasks and Evaluation in Civic Education]. Schwalbach/Taunus, 23–40.
- Massing, Peter. 2006. Bildungsstandards als Diagnoseinstrument für den Politikunterricht [Standards of Education as an Instrument for Diagnosing Political Education]. In: Richter, Dagmar; Schelle, Carla, eds. *Politikunterricht evaluieren. Ein Leitfaden zur fachdidaktischen Unterrichtsanalyse* [Evaluation Political Education. Compendium for the Subject-Didactic Analysis of Education]. Baltmannsweiler, 141–160.
- May, Michael. 2008. *Demokratielernen oder Politiklernen?* [Citizenship Education or Civic Education] Schwalbach/Taunus.
- McNeil, Linda. 2001. *Contradictions of School Reform. Educational Costs of Standardized Testing*. New York.
- Mittelstraß, Jürgen. 2003. Transdisziplinarität – wissenschaftliche Zukunft und institutionelle Wirklichkeit [Transdisciplinarity – Scientific Future and Institutional Reality]. Konstanz.
- Moran, Joe. 2002. *Interdisciplinarity*. London.
- Murphy, James Bernard. 2007. Against Civic Education in Public Schools. In: *International Journal of Public Administration*, Vol. 30, 651–670.
- Nichols, Sharon L; Berliner, David C. 2007. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge.
- Noelle-Neumann, Elisabeth; Köcher, Renate. 2002. *Allensbacher Jahrbuch der Demoskopie 1998-2002* [Allensbacher Annual of Opinion Poll 1998 – 2002]. München.
- Oancea, Alis; Pring, Richard. 2009. The Importance of Being Thorough: On Systematic Accumulations of 'What Works' in Education Research. In: Bridges, David; Smeyers, Paul; Smith Richard, eds. *Evidence-Based Education Policy. What Evidence? What Basis? Whose Policy?* Malden MA, 11–35.

- Oesterreich, Detlef. 2002. Politische Bildung von 14-Jährigen in Deutschland. Studie aus dem Projekt Civic Education [Civic Education of 14-years-old Youths in Germany. Studies of the Project Civic Education]. Opladen.
- Reinhardt, Sibylle. 2007. Politik-Didaktik. Praxishandbuch für die Sekundarstufe I und II [Political-Didactics. Practical Handbook for Secondary Education I and II]. Berlin.
- Rippel, Susanne. 2004. Umweltbewusstsein und Umweltverhalten. Ein empirischer Theorienvergleich aus kulturvergleichender Perspektive [Environmental Awareness and Environmental Behavior. An Empirical Comparison of Theories from a Cross-Cultural Perspective]. In: http://www.qucosa.de/fileadmin/data/qucosa/documents/6192/habil_rippel.pdf
- Sacks, Peter. 2000. Standardized minds: the high price of America's .testing culture and what we can do to change it. New York.
- Schattschneider, Jessica, eds. 2007. Domänenspezifische Diagnostik. Wissenschaftliche Beiträge für die politische Bildung [Domain-Specific Diagnostics. Scientific Contributions to Civic Education]. Schwalbach/Taunus.
- Schneekloth, Ulrich. 2010. Jugend und Politik. Aktuelle Entwicklungen und Perspektiven [Youth und Politics. Current Trends and Perspectives]. In: Shell Deutschland Holding, eds. Jugend 2010. Eine pragmatische Generation behauptet sich [A Pragmatic Generation comes to stay]. Frankfurt am Main, 129–164.
- Schulz, Wolfram; Ainley, John; Fraillon, Julian; Kerr, David, Losito Bruno. 2010a. Initial Findings from the IEA International Civic and Citizenship Education Study. (http://www.iea.nl/fileadmin/user_upload/ICCS/ICCS_2010_Initial_Findings.pdf).
- Schulz, Wolfram; Ainley, John; Fraillon, Julian; Kerr, David, Losito Bruno. 2010b. ICCS 2009 International Report: Civic Knowledge, Attitudes, and Engagement among lower-secondary School Students in 38 Countries. Amsterdam. (http://www.iea.nl/fileadmin/user_upload/ICCS/ICCS_InternationalReport.pdf).
- Schulz, Wolfram. 2009. Questionnaire construct Validation in the International Civic and Citizenship Education Study. (http://www.ierinstitute.org/IERI_Monograph_Volume_02_Chapter_05.pdf).
- Sukopp, Thomas. 2010. Interdisziplinarität und Transdisziplinarität. Definition und Konzepte [Interdisciplinarity and Transdisciplinarity. Definition and Concepts]. In: Jungert, Michael; Romfeld, Elsa; Sukopp, Thomas; Voigt, Uwe, eds. Interdisziplinarität. Theorie, Praxis, Probleme [Interdisciplinarity. Theory, Practice, Problems]. Darmstadt, 13–12.
- Van de Vijver F. J. R. 1998. Towards a Theory of Bias and Equivalence. In: Harkness, J.A., ed. Cross-Cultural Survey Equivalence, Zuma-Nachrichten, Spezial Band 3. Mannheim: ZUMA, 41–66.
- Van der Vijver, F. J. R.; Hambleton, R. K. 1996. Translating Tests: Some Practical Guidelines. In: European Psychologist, 1, 89–99.
- Verbeet, Markus; Leffers, Jochen 2009. Studenten-Pisa. Frauen patzen, 26 Neunmalkluge triumphieren [Student-Pisa. Women Blunder, 26 Smart-Alecks Triumph]. In: Spiegel Online (www.spiegel.de/unispiegel/studium/0,1518,625140,00.html) published 18.05.2009.
- Wolak, Jennifer; Mc Devitt, Michael. 2010. The Roots of the Gender Gap in Political Knowledge in Adolescence. In: Political Behavior, 6 October 2010, 1–29.
- Weinbrenner, Peter. 1997. Politische Urteilsbildung als Ziel und Inhalt des Politikunterrichts [Political Judgement as a Purpose and Substance in Political Education]. In: Bundeszentrale für politische Bildung [Federal Agency for Civic Education].